

IN THE CLAIMS:

The following is a complete list of the claims now pending; this listing replaces all earlier versions and listings of the claims.

5/27
1. (Currently Amended) A document type definition generating method;
~~comprising, in for generating a document type definition of a structured document provided with~~
~~a tag tags, each tag having an element name in for each document element, said method~~
comprising:

a physical structure judging step₁ of judging a similarity between
~~physical structure of each~~ structures of each of the document element elements in the structured
document;

a semantic structure judging step₂ of judging a similarity between
~~semantic structure of said each document element~~ structures of each of the tags by comparing the
form of each tagged element; and

a document type definition generating step₃ of generating document
~~type definition to define appearance state of the document element in said structured document~~
judging a similarity of the tags based on judgment results of said physical structure judging step
and said semantic structure judging step₂, and generating the document type definition which
unifies the element names of the similar tags.

2. (Currently Amended) ~~The~~ A document type definition generating
method according to claim 1, wherein said physical structure judging step comprises judging the

physical structure of the document ~~element~~ elements based on an ~~indentation~~ indentation or a blank line.

3. (Currently Amended) ~~The~~ A document type definition generating method according to claim 2, wherein, when the physical structure of the document element is judged based on ~~said-indentation~~ the indentation, the judging is performed by excluding the ~~indentation~~ indentation which represents quotation.

4. (Currently Amended) ~~The~~ A document type definition generating method according to claim 2, wherein, when the physical structure of the document element is judged based on ~~said~~ the blank line, the judging is performed by excluding the blank line from ~~a~~ the document in which a description is made by constantly placing every predetermined number of blank lines.

5. (Original) ~~The~~ A document type definition generating method according to claim 1, wherein said physical structure judging step comprises judging the physical structure of the document element based on a positional ~~relation~~ relationship of the tags surrounding the document element.

6. (Currently Amended) ~~The~~ A document type definition generating method according to claim 1, wherein said semantic structure judging step comprises referring to

a semantic information database to judge the semantic structure of the document element based on connection of words and phrases ~~connection~~ in a the document and word types.

7. (Original) ~~The~~ A document type definition generating method according to claim 1, wherein said semantic structure judging step comprises judging the semantic structure of the document element based on a meaning represented by the tags surrounding the document element.

8. (Currently Amended) ~~The~~ A document type definition generating method according to claim 1, wherein said document type definition generating step comprises a redundancy removing step of, when the physical structure and the semantic structure of a plurality of document elements having the tags different in element name are similar, regarding the document elements as being of the same type and excluding one element name from a document type definition generating object based on the judgment results of said physical structure judging step and said semantic structure judging step.

9. (Currently Amended) ~~The~~ A document type definition generating method according to claim 8, wherein said redundancy removing step comprises obtaining similarity degrees concerning agreement degrees of the physical structure and the semantic structure between the document elements having the tags different in element name, and regarding the document elements as being of the same type when a general similarity value calculated from the similarity degrees is equal to or more than a predetermined threshold value.

10. (Currently Amended) ~~The~~ A document type definition generating method according to claim 1, wherein said document type definition generating step comprises a title changing step of, when the physical structure and the semantic structure of a plurality of document elements having ~~the~~ tags with the same element name are different, regarding the document elements as being of different types and changing one element name based on the judgment results of said physical structure judging step and said semantic structure judging step.

11. (Original) ~~The~~ A document type definition generating method according to claim 1, wherein said document type definition generating step comprises analyzing words and phrases present between a start tag and an end tag having the same title, obtaining information to be included between the tags, and generating the document type definition based on the information.

12. (Currently Amended) A document type definition generating apparatus ~~comprising: in~~ for generating a document type definition of a structured document provided with ~~a tag~~ tags, each tag having an element name in for each document element, said apparatus comprising:

physical structure judging means for judging a similarity between ~~physical structure of said each~~ structures of each of the document ~~element~~ elements in the structured document;

semantic structure judging means for judging a similarity between
~~semantic structure of said each document element~~ structures of each of the tags by comparing the
form of each tagged element; and

document type definition generating means for ~~generating document~~
~~type definition to define appearance state of the document element in said structured document~~
judging a similarity of the tags based on judgment results of said physical structure judging
means and said semantic structure judging means, and generating the document type definition
which unifies the element names of the similar tags.

13. (Currently Amended) The A document type definition generating
apparatus according to claim 12, wherein said physical structure judging means judges the
physical structure of the document ~~element~~ elements based on an ~~indentation~~ indentation or a blank
line.

14. (Currently Amended) The A document type definition generating
apparatus according to claim 13, wherein said physical structure judging means judges the
physical structure of the document element based on ~~said indentation~~ the indentation by excluding
the ~~indentation~~ indentation which represents quotation.

15. (Currently Amended) The A document type definition generating
apparatus according to claim 13, wherein said physical structure judging means judges the
physical structure of the document element based on ~~said~~ the blank lines by excluding the blank

lines from a document in which description is made by constantly placing every predetermined number of blank lines.

16. (Original) ~~The~~ A document type definition generating apparatus according to claim 12, wherein said physical structure judging means judges the physical structure of the document element based on a positional relation of the tags surrounding the document element.

sub
B7
17. (Currently Amended) ~~The~~ A document type definition generating apparatus according to claim 12, wherein said semantic structure judging means refers to a semantic information database to judge the semantic structure of the document element based on connection of words and phrases ~~connection~~ in a the document and word types.

18. (Original) ~~The~~ A document type definition generating apparatus according to claim 12, wherein said semantic structure judging means judges the semantic structure of the document element based on a meaning represented by the tags surrounding the document element.

19. (Currently Amended) ~~The~~ A document type definition generating apparatus according to claim 12, wherein said document type definition generating means comprises redundancy removing means for, when the physical structure and the semantic structure of a plurality of document elements having the tags different in element name are

similar, regarding the document elements as being of the same type and excluding one element name from a document type definition generating object based on the judgment results of said physical structure judging means and said semantic structure judging means.

20. (Currently Amended) ~~The~~ A document type definition generating apparatus according to claim 19, wherein said redundancy removing means obtains similarity degrees concerning agreement degrees of the physical structure and the semantic structure between the document elements having ~~the~~ tags different in element name, and regards the document elements as being of the same type when a general similarity value calculated from the similarity degrees is equal to or more than a predetermined threshold value.

21. (Currently Amended) ~~The~~ A document type definition generating apparatus according to claim 12, wherein said document type definition generating means comprises title changing means for, when the physical structure and the semantic structure of a plurality of document elements having ~~the~~ tags with the same element name are different, regarding the document elements as being of different types and changing one element name based on the judgment results of said physical structure judging means and said semantic structure judging means.

22. (Original) ~~The~~ A document type definition generating apparatus according to claim 12, wherein ~~said~~ document type definition generating means analyzes words

and phrases present between a start tag and an end tag having the same title, obtains information to be

included between the tags, and generates the document type definition based on the information.

23. (Currently Amended) A computer-readable storage medium storing a document type definition generating program for controlling a computer to ~~perform~~ execute a document type definition generation method for generating document type definition of a structured document provided with tags, each tag having an element name for each document element, said program comprising ~~codes for causing the computer to perform:~~

~~in a structured document provided with a tag having an element name in each document element; code for a physical structure judging step, of judging a similarity between physical structure of each structures of each of the document element elements in the structured document;~~

~~code for a semantic structure judging step, of judging a similarity between semantic structure of said each document element structures of each of the tags by comparing the form of each tagged element; and~~

~~code for a document type definition generating step, of generating document type definition to define appearance state of the document element in said structured document judging a similarity of the tags based on judgment results of said physical structure judging step and said semantic structure judging step, and generating the document type definition which unifies the element names of the similar tags.~~